

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:17:16 ; Search time 52.57 Seconds  
(without alignments)  
122.626 Million cell updates/sec

Title: US-09-331-631a-25\_COPY\_31\_85

Perfect score: 315  
Sequence: 1 ENPKHNKCLQSCNSERDSYR.....EECEGEIYPRRPQHPER 55

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_15:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175	55.6	62	10 Q9M402	Q9M402 glycine max
2	131	41.6	544	10 Q22120	Q22120 glycine max
3	69.5	22.1	556	5 Q96456	Q96456 plasmodium
4	69.5	22.1	556	5 Q9TVE0	Q9TVE0 plasmodium
5	67.5	21.4	719	5 Q9M4B8	Q9M4B8 drosophila
6	67	21.3	593	10 Q9SBW4	Q9SBW4 juglans reg
7	66.5	21.1	3680	5 Q9VR08	Q9VR08 drosophila
8	66	21.0	963	10 Q9M8T1	Q9M8T1 arabidopsis
9	65.5	20.8	510	5 Q94727	Q94727 plasmodium
10	65.5	20.8	774	3 Q42918	Q42918 schizosacch
11	64.5	20.5	340	11 Q08641	Q08641 mus musculu
12	64.5	20.5	704	13 Q90865	Q90865 gallus gall
13	63.5	20.2	414	5 Q20155	Q20155 caenorhabdi
14	63	20.0	663	5 Q62446	Q62446 streptococc
15	62.5	19.8	519	2 Q9KR20	Q9KR20 caenorhabdi
16	62.5	19.8	3767	5 Q9UA13	Q9UA13 caenorhabdi
17	62	19.7	190	10 Q9SBH3	Q9SBH3 arabidopsis
18	62	19.7	190	10 Q9SBG9	Q9SBG9 arabidopsis
19	62	19.7	238	10 Q81922	Q81922 capsicum an

20	62	19.7	244	10 Q9SUD9	Q9SUD9 arabidopsis
21	62	19.7	255	10 Q39081	Q39081 arabidopsis
22	62	19.7	462	10 Q39249	Q39249 arabidopsis
23	62	19.7	607	5 Q9V6S7	Q9V6S7 drosophila
24	62	19.7	807	13 Q91A01	Q91A01 ictalurus p
25	61.5	19.5	191	10 Q9SBH2	Q9SBH2 arabidopsis
26	61	19.4	147	5 Q23199	Q23199 caenorhabdi
27	61	19.4	190	10 Q9SBH5	Q9SBH5 arabidopsis
28	61	19.4	472	10 Q9SM43	Q9SM43 spiraea ol
29	61	19.4	757	5 Q9VZF2	Q9VZF2 drosophila
30	61	19.4	1952	3 Q9P6T1	Q9P6T1 neurospora
31	60.5	19.2	600	4 Q9S402	Q9S402 homo sapien
32	60.5	19.2	2277	5 Q9U0G5	Q9U0G5 mus musculu
33	60	19.0	68	11 Q70562	Q70562 mus musculu
34	60	19.0	190	10 Q9SBH0	Q9SBH0 mus musculu
35	60	19.0	1132	4 P78418	P78418 homo sapien
36	59.5	18.9	1101	4 Q9NXY9	Q9NXY9 homo sapien
37	59	18.7	76	11 Q70536	Q70536 mus musculu
38	59	18.7	190	10 Q9SBH6	Q9SBH6 arabidopsis
39	59	18.7	190	10 Q9SBH4	Q9SBH4 arabidopsis
40	59	18.7	190	10 Q9SBH1	Q9SBH1 arabidopsis
41	59	18.7	190	10 Q9SBG8	Q9SBG8 arabidopsis
42	59	18.7	190	10 Q9S780	Q9S780 arabidopsis
43	59	18.7	190	10 Q9S779	Q9S779 arabidopsis
44	59	18.7	199	5 P91512	P91512 caenorhabdi
45	59	18.7	281	5 Q23778	Q23778 chironomus

#### ALIGNMENTS

RESULT	1				
ID	Q9M402	PRELIMINARY;	PRT;	62 AA.	
AC	Q9M402:				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DE	01-OCT-2000 (TREMBLrel. 15, Last annotation update)				
DE	ALPHA'-TYPE BETA-CONGLYCININ (FRAGMENT).				
GN	BCSP.				
OS	Glycine max (Soybean).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;				
OC	Fabales; Fabaceae; Papilionoideae; Glycine.				
OX	NCBI_TaxID=3847;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Morasutti C., Marchetti S., De Amicis F., Zaina G., Zanetti M.;				
RT	"Expression and purification of the antimicrobial peptide SMAP-29 from				
RT	tobacco transgenic plants by ScevMA-intein system."				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	De Amicis F.;				
RL	Thesis (1998), University of Udine, DPVTA.				
DR	EMBL: AJ276118; CAB83040.1; -				
FT	NON FR	62			
FT	NON FR	62			
SQ	SEQUENCE	62 AA;	7158 MW;	10690053F7700443 CRC64;	
Query Match					
Best Local Similarity 55.6%; Score 175; DB 10; Length 62;					
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1	ENPKHNKCLQSCNSERDSYRNQACHRCNLK	32		
DB	31	EIPKHNKCLQSCNSERDSYRNQACHRCNLK	62		
RESULT	2				
ID	Q22120	PRELIMINARY;	PRT;	544 AA.	
AC	Q22120:				

DT 01-JAN-1998 (TREMBLrel.. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel.. 05, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel.. 15, Last annotation update)  
 DE ALPHA SUBUNIT OF BETA CONGLICININ.  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Glycyne.  
 OX NCBI\_TaxID=38447;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. WASESUZUNARI;  
 RA Matuyama N., Katsube T., Wada Y., Utsumi S.;  
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB008678; BAA23360.1; -.  
 DR HSSP; P50477; ICAU.  
 DR MENDEL; 25074; Glyma;1188;25074.  
 DR INTERPRO: IPR001113; -.  
 DR PfAM: PF00546; Seedstore\_7s; 1.  
 DR PRODOM: PD081059; -.  
 SO SEQUENCE 544 AA; 63296 MW; 04BD243929A9C9AB CRC64;

Query Match	41.6%	Score	131	DB	10	Length	544
Best Local Similarity	100.0%	Pred. No.	4	5e-08			
Matches	23	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

```
QY      33 VEKECEGEIPRRPRPQHPER 55
          |||||
Db      2  VEKECEGEIPRRPRPQHPER 24
```

RESULT	3			
096456	ID	096456	PRELIMINARY;	PRT; 556 AA.
AC	096456;			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, last annotation update)			
DE	SPOROZOITE SURFACE PROTEIN 2.			
GN	SSP2.			
OS	Plasmodium vivax.			
OC	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5855;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=SL. I;			
RA	Rogers W.O., Gowda K., Sacti J., Hoffman S.L.;			
RT	"Construction and Immunogenicity of DNA vaccine plasmids encoding four			
RT	Plasmodium vivax candidate vaccine antigens.";			
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF063136; AAC97484.1; -			
DR	INTERPRO; IPR000884; -			
DR	INTERPRO; IPR002035; -			
DR	PFAM; PF00090; tsp_1; 1.			
DR	PFAM; PF00092; vwa; 1.			
DR	PRINTS; PR00453; VMFODMAIN.			
FT	VARIANT 134 134 D -> E.			
FT	VARIANT 160 160 V -> I.			
FT	VARIANT 166 166 R -> T.			
FT	VARIANT 172 172 N -> K.			
FT	VARIANT 268 268 K -> G.			
FT	VARIANT 352 352 D -> E.			
FT	VARIANT 366 366 T -> A.			
FT	VARIANT 391 391 I -> R.			
FT	VARIANT 464 464 E -> Q.			
FT	VARIANT 494 494 T -> I.			
Q	SEQUENCE 556 AA; 61208 MW; 530B5A0C86EBF3A7 CRC64;			

Query Match	22.18;	Score 69.5;	DB 5;	Length 556;
Best Local Similarity	33.38;	Pred. No. 1.2;		
Matches 18;	Conservative 4;	Mismatches 23;	Indels 9;	Gaps 2

```

QY      8  CLOSCNSERDSTYRNQACHARC�LLKVEKECECE-----LPRKRPQPHE 54
          |  :  |  :  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    249  CSVTCGRGTHSRSPSLHEKCTHHM--SECCEGECVPVEBPPLVPAPLPYPE 300

```

RESULT	4			
O9TVFO			PRT;	556 AA.
ID	O9TVFO	PRELIMINARY;		
AC	O9TVFO			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-JUN-2000 (TrEMBLrel. 14, last annotation update)			
DE	SPOROZOITE SURFACE PROTEIN 2.			
CN	spp2.			
OS	Plasmodium vivax.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5855;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SAL I;			
RX	MEDLINE=99389346; PubMed=10462250;			
RT	Rogers W.O., Gowda K., Hoffman S.L.;			
RT	"Construction and immunogenicity of DNA vaccine plasmids encoding four			
RL	vaccine 17:3136-3144(1999)."			
DR	EMBL; AF063137; AAC97485.1; -.			
DR	INTERPRO; IPR000884; -.			
DR	INTERPRO; IPR002035; -.			
DR	PFAM; PFD0090; tsp_1; 1.			
DR	PEAM; PF00092; vwa; 1.			
DR	PRINTS; PR00453; VWFADOMAIN.			
FT	VARIANT	134	D -> E.	
FT	VARIANT	160	V -> I.	
FT	VARIANT	166	R -> T.	
FT	VARIANT	172	N -> K.	
FT	VARIANT	268	K -> G.	
FT	VARIANT	352	D -> E.	
FT	VARIANT	391	M -> R.	
FT	VARIANT	464	E -> Q.	
SO	SEQUENCE	556 AA;	61208 MM;	7FAFE294B3BE3ZAA7 CRC64;

Query Match	22.1%	Score 69.5;	DB 5;	Length 556;
Best Local Similarity	33.3%	Pred. No. 1.2;		
Matches	18;	Conservative	4;	Mismatches 23;
			Indels	9;
			Gaps	2

QY 8 CLQCSNSERDSTYNQACHARNLKEVEKECEEGE-----IPRRPRQPHE 54  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 249 CSVTCGRGTHSRSDSLHEKCTPHMV--SECEEGEPVEPEPLPVAPRLPTVPE 300

RESULT	5		
Q9M4B8			
ID	Q9M4B8	PRELIMINARY;	PRT; 719 AA.
AC	Q9M4B8;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)		
DE	CG3239 PROTEIN.		
GN	CG3239.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY.		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gccayne J.D.,		
RA	Amaratilaks P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		

RA Sutton G.G., Mortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
RA Abtil J.F., Agbayanl A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bowen B.P., Bhandari D., Bolshakov S.,  
RA Borckova D., Botchan M.R., Bouck J., Brookstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Duda J.K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.Z., Gelbart W.M., Glasser K.,  
RA Glaser A., Gong F.H., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,  
RA Jatalai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., McCarthy C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong E.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gladys R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL, AE003434, AAP6038.1; -  
DR FlyBase, FBgn0029769; CG3239.  
DR INTERPRO, IPR000130; -  
DR INTERPRO, IPR000718; -  
DR Pfam, PF01431; Peptidase\_M13; 1.  
DR PRINTS, PRO0786; NEPRILYSIN.  
DR PROSITE, PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SOURCE 71 AA: 83325 MW: 663DE7BA3802C870 CRC64;

```

Query Match 21.4%: Score 67.5; DB 5; Length 719;
Best Local Similarity 27.1%: Pred. No. 2.7;
Matches 23; Conservative 8; Mismatches 17; Indels 37; Gaps 5;

OY 3 PRHNCC-----LQSCNERSDS-----YRNOACHA--RCNLLKVEKE 36
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 PCYNCPAPRPVYCKRSPSQICVNSKRNROCKVLANSCQLRNQHCSPRNMLLTDR 149
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 37 ECCE---GEIP-----RPRPR 50
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 RCGQLQLODKRPQNCIRPVPRPRPT 174

RESULT 6
O9SEW4
ID O9SEW4 PRELIMINARY; PRT: 593 AA.
O9SEW4;
DT 01-MAY-2000 (TREMBLrel, 13, Created)
DT 01-MAY-2000 (TREMBLrel, 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel, 15, Last annotation update)
DE VICLIN-LIKE PROTEIN PRECURSOR (FRAGMENT).
OS Eugenia regia (English walnut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fagales; Juglandaceae; Juglans.
OX NCBI_TextID=51240;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-CV. SUNJIAN; TISSUE-SOMATIC EMRO LINE;  
RA Teuber S.S., Jarvis K.C., Peterson W.R., Dandekar A.V., Ansari A.A.;  
RT "Identification and cloning of a cDNA encoding a vicillin-like protein,  
Tug1-2, from English walnut kernel (Juglans regia): a major food  
allergen.";  
RT Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF066055; AAF18269.1; -.  
DR HSSP; P02853; 2PHL. .  
DR INTERPRO; IPR001113. -.  
DR PFWA; PF00546; Seedscore-7s; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE 593 AA; 69990 MW; 9BA127E19B18C0D8 CRC64;

```

Query Match      21.3%; Score 67; DB 10; length 593;
Best Local Similarity 28.3%; Pred. No. 2.6;
Matches 17; Conservative 10; Mismatches 23; Indels 10; Gaps 3

QY      1 ENP-----KHKKKQSCNSESDSTRNQACCHARNLKAYKECEGCEGTIPRPPRPPQHEH 55
      ::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db       67 QNPDEPRQRQQCCQCCCRGRGGRGGQTLGRRCCE---QRRQQRERERGRGDR--QDPQ 121

```

RESULT	7			
ID	09VR08	PRELIMINARY:	PRF:	3680 AA.
AC	09VR08;			
DT	01-MAY-2000 (TREMBlrel, 13, Created)			
DT	01-MAY-2000 (TREMBlrel, 13, Last sequence update)			
DT	01-OCT-2000 (TREMBlrel, 15, Last annotation update)			
DE	CG15637 PROTEIN.			
GN				
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephyroidea; Drosophilidae; Drosophila.			
NCBI	NCBI_TaxId=7227;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sulton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,			
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mkios G.L.G.,			
RA	Abriil J.F., Adayyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintshkov S.,			
RA	Borkus K.C., Botchan M.R., Bouck H., Brokstein P., Brotlier P.,			
RA	Cherry J.M., Cawley S., Dahlke S., Daltveit A., Davidson L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,			
RA	Doudson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz S., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Gang N.S., Galhart W.M., Glasser K.,			
RA	Godet A., Gong F., Gorrell I.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mallet B., McIntosh T.C., McLeod M.P., McKernson D.,			
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Switz R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Weissman D.A., Weinstock G.M., Weissbach J.,			

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003576; AAF51000.1; -.  
 DR HSSP: P35535; 1EMN.  
 DR FLYBASE: FBgn0031612; CG15637.  
 DR INTERPRO: IPR000152; -.  
 DR INTERPRO: IPR000561; -.  
 DR INTERPRO: IPR001881; -.  
 DR PFAM: PF00008; EGF\_28.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 18.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 18.  
 DR PROSITE: PS01187; EGF\_CA; 17.  
 SO SEQUENCE 3680 AA; 388637 MW; E5E972E1A3479EFF CRC64;

Query Match 21.1%; Score 66.5; DB 5; Length 3680;  
 Best Local Similarity 27.4%; Pred. No. 15;  
 Matches 17; Conservative 8; Mismatches 20; Indels 17; Gaps 3;

OY 5 HNCYCLOSCN-----SEDSYRNQAC-----HARCNLKVEKE-ECGEGELPRPR 47  
 DB 2818 HNCYVCGCHVDGDCSASESCRNDCVNPCLNCPGPMACVSNIHRASCSLESMPVNPRT 2877  
 OY 48 PR 49  
 DB 2878 PQ 2879

RESULT 8 PRELIMINARY: PRT: 963 AA.  
 ID O9M8T1  
 AC O9M8T1:  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE F13E7.16 PROTEIN.  
 GN F13E7.16.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
 RA Rongling C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome III BAC F13E7 genomic sequence.";  
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC018363; AAF26970.1; -.  
 SO SEQUENCE 963 AA; 105391 MW; D204IDC304280339 CRC64;

Query Match 21.0%; Score 66; DB 10; Length 963;  
 Best Local Similarity 32.3%; Pred. No. 5.3;  
 Matches 21; Conservative 3; Mismatches 13; Indels 28; Gaps 3;

OY 1 ENPKHNKCL-----QSCNSERDSYRNQACHARCNLKVEKE-----EC 38  
 DB 832 ENPNSNKSLLTDRTLSMQSCMKEDAKKEKACSG-----TEKNMFVSYSGEGVDVET 885  
 OY 39 EGGEL 43  
 DB 886 EGGEL 890

RESULT 9

O94727  
 ID O94727 PRELIMINARY: PRT: 510 AA.  
 AC O94727:  
 DT 01-FEB-1997 (Tremblrel. 02, Created)  
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE THROMBOSPONDIN-RELATED ANONYMOUS PROTEIN (FRAGMENT).  
 GN TRAP.  
 OS Plasmidium vivax.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.  
 OX NCBI\_TaxID=5855;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SAL-I;  
 RX MEDLINE=97193932; PubMed=9041517;  
 RA Templeton T.J., Kaslow D.C.;  
 RT "Cloning and cross-species comparison of the thrombospondin-related  
 RT anonymous protein (TRAP) gene from *Plasmidium knowlesi*, *Plasmidium*  
 RT vivax and *Plasmidium gallinaceum*.";  
 RL EMBL: U64901; AAC47463.1; -.  
 DR INTERPRO: IPR000884; -.  
 DR INTERPRO: IPR002035; -.  
 DR PFAM: PF00090; tSP\_1; 1.  
 DR PFAM: PF00092; vwa; 1.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 FT NON\_TER 1 1  
 FT NON\_TER 510 510  
 SO SEQUENCE 510 AA; 55601 MW; 72687C7BAFDFA1D CRC64;

Query Match 20.8%; Score 65.5; DB 5; Length 510;  
 Best Local Similarity 33.3%; Pred. No. 3.5;  
 Matches 18; Conservative 3; Mismatches 24; Indels 9; Gaps 2;

OY 8 CLOSCSERDSYRNQACHARCNLKVEKE-----IPRRPRPQHE 54  
 DB 210 CSVTCGRGTHSRSPSLHRCCTTHMV--SECEGEGCVPEPLVPAPLPYPE 261

RESULT 10  
 ID O42918 PRELIMINARY: PRT: 774 AA.  
 AC O42918:  
 DT 01-JAN-1999 (Tremblrel. 09, Created)  
 DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE PUTATIVE ALPHA-AMYLASE C16A3.13 PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-  
 DE GLUCAN GLUCANOHYDROLASE).  
 GN SPCH16A3.13.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
 OC Schizosaccharomycetaceae; Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Beck A., Reinhardt R.;  
 RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC  
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.  
 CC -1- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: TO THE FAMILY 13 OF GLUCOSYL HYDROLASES, ALSO KNOWN AS  
 CC THE ALPHA-AMYLASE FAMILY.  
 DR EMBL: A1021748; CA16864.1; -.  
 DR HSSP: P10529; 7TAA.  
 DR INTERPRO: IPR000461; -.  
 DR PFAM: PF00128; alpha-amylase; 1.  
 KW Hypothetical protein; Hydrolase; glycosidase; Carbohydrate metabolism;  
 KW Calcium; Glycoprotein; Signal.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 774 PUTATIVE ALPHA-AMYLASE C16A3.13.

```

FT ACT_SITE 461 461 BY SIMILARITY.
FT ACT_SITE 553 553 BY SIMILARITY.
FT DISULFID 52 60 BY SIMILARITY.
FT DISULFID 229 314 BY SIMILARITY.
FT DISULFID 495 539 BY SIMILARITY.
FT DISULFID 690 724 BY SIMILARITY.
FT CARBOHYD 143 143 POTENTIAL.
FT CARBOHYD 187 187 POTENTIAL.
FT CARBOHYD 214 214 POTENTIAL.
FT CARBOHYD 243 243 POTENTIAL.
FT CARBOHYD 272 272 POTENTIAL.
FT CARBOHYD 299 299 POTENTIAL.
FT CARBOHYD 326 326 POTENTIAL.
FT CARBOHYD 355 355 POTENTIAL.
FT CARBOHYD 382 382 POTENTIAL.
FT CARBOHYD 409 409 POTENTIAL.
SQ SEQUENCE 774 AA: 89459 MW: 92CEAE45664269B CRC64;

```

```

Query Match 20.8%; Score 65.5; DB 3; Length 774;
Best Local Similarity 28.1%; Pred. No. 5;
Matches 16; Conservative 8; Mismatches 20; Indels 13; Gaps 3;

```

```

OY 6 NKLQSCNSRSDYRNQACHARCLLVE-----KECEEGEIPRRP-----RPOH 52
Db 167 HKCEHSCSHHDFPRPVH--NGTRDHRPKWKEHCHHGKPRPVPHNGTKRPDH 220

```

RESULT 11

```

ID 008641 PRELIMINARY; PRT; 340 AA.

```

```

AC 008641;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE RAY PROTEIN.
GN SH3YL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL6; TISSUE=SKIN;
RA Aoki N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D85926; BAA19679.1; -.
DR HSSP: P29354; 1GFC.
DR MCD; MGI:1346118; Sh3yl1.
DR INTERPRO: IPR001452; -.
DR PFAM: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 340 AA: 37028 MW: 3FE12C0A8530C74E CRC64;

```

```

Query Match 20.5%; Score 64.5; DB 11; Length 340;
Best Local Similarity 38.6%; Pred. No. 3.2;
Matches 17; Conservative 8; Mismatches 14; Indels 5; Gaps 3;

```

```

OY 13 NSESDYRNQACHARCLLVEKECEEGEIPRRP--RPPPOH 53
Db 208 NSFTEKYETEG--QRINLKRVAREORAKELPPKPSRPOPAH 249

```

RESULT 12

```

ID 090865 PRELIMINARY; PRT; 704 AA.

```

```

AC 090865;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HEPATOCYTE GROWTH FACTOR-LIKE/MACROPHAGE STIMULATING PROTEIN.

```

```

GN HGFI/MSF.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=96029010; PubMed=7554499;
RA Thery C., Sharpe M.J., Bally S.J., Stern C.D., Gherardi E.;
RT "Expression of HGF/SF, HGFI/MSF, and c-met suggests new functions
during early chick development.";
RL Dev. Genet. 17:90-101(1995).
DR EMBL: X84043; CA58862.1; -.
DR HSSP: P00747; 1CEA.
DR MEROPS: S01.977; -.
DR INTERPRO: IPR000001; -.
DR INTERPRO: IPR001254; -.
DR INTERPRO: IPR001314; -.
DR INTERPRO: IPR003014; -.
DR PFAM: PF000024; PAN; 1.
DR PFAM: PF000051; kingle; 4.
DR PFAM: PF000089; trypsin; 1.
DR PRINTS: PR00018; KRINGLE.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00021; KRINGLE_1; UNKNOWN_4.
DR PROSITE: PS50070; KRINGLE_2; 4.
SQ SEQUENCE 704 AA: 79341 MW: CAB0D8CC41367C37 CRC64;

```

```

Query Match 20.5%; Score 64.5; DB 13; Length 704;
Best Local Similarity 23.8%; Pred. No. 6.1;
Matches 15; Conservative 7; Mismatches 24; Indels 17; Gaps 3;

```

```

OY 2 NPKHNKC-----LQSCNSRSDYRNQACHARCLLVEKECEEGEIPRRPPOH 52
Db 173 NVRHQSGIKKCDAYCMTNGE--DYRGFVDHTESG-----TECORWLDQPHKPHYH 224

```

```

OY 53 PER 55
Db 225 PDK 227

```

RESULT 13

```

ID 020155 PRELIMINARY; PRT; 414 AA.

```

```

AC 020155;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE F38B7.1 PROTEIN.
GN F38B7.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RL Leonard N.;
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC OF FORM A.
DR EMBL: Z74033; CA98475.1; -.
DR EMBL: Z74033; CA98476.1; -.
DR INTERPRO: IPR000571; -.
DR PFAM: PF00642; zf-CCH; 2.
DR Hypothetical protein; Alternative splicing.
FT VARSPLIC 1 38 MISSING (IN ISOFORM B).
FT VARSPLIC 39 52 MISSING (IN ISOFORM B).
SQ SEQUENCE 414 AA: 44308 MW: D8B29D29A9E39583 CRC64;

```

```

Query Match          20.2%; Score 63.5; DB 5; Length 414;
Best Local Similarity 35.7%; Pred. No. 5.1;
Matches 20; Conservative 7; Mismatches 14; Indels 15; Gaps 5;

QY 1 ENPKHNCLOSCNSERDSYRNQACHARCNLKVEKEBCE--ESEEIPRPPRPOHPE 54
      ||| | : | | | | | | | | | | | | | | | | | | | | | | | |
DB 194 KNRKLYK-TELGRSMD-----HGRCNY----GRCQYAHGELER-RVPRHPK 236

RESULT 14
O62446 PRELIMINARY; PRT; 663 AA.
ID O62446;
AC O62446;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE Y43F4A.1 PROTEIN.
GN Y43F4A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Koopa A., Saunders D., Showkhen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
DR EMBL: 299271; CAB16471.1; -.
DR INTERPRO: IPR000130; -.
DR INTERPRO: IPR001577; -.
DR PFAM: PF01457; Peptidase_M8; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 663 AA; 76049 MW; 814C31728B5ACEA CRC64;

Query Match          20.0%; Score 63; DB 5; Length 663;
Best Local Similarity 35.9%; Pred. No. 8.8;
Matches 14; Conservative 4; Mismatches 17; Indels 4; Gaps 1;

QY 7 KCLQSCNSERDSYRNQACHARCNLKVEKEBCEGEP 45
      ||| | : | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 KCISFYKYGKGRNVACDKGR---ERTTCGEADIPR 155

RESULT 15
O9KK20 PRELIMINARY; PRT; 519 AA.
ID O9KK20;
AC O9KK20;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE SURFACE PROTEIN PSPC.
GN PSPC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.

```

```

OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=699;
RA Tannelli F., Oggioni M.R., Spinoza M.R., Pozzi G.;
RT "Allelic variants of surface protein PspC (SpsA, CpaA, PbcA) of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF154036; AAF73808.1; -.
SQ SEQUENCE 519 AA; 58731 MW; 7E1879295825AFD1E CRC64;

```

```

Query Match          19.8%; Score 62.5; DB 2; Length 519;
Best Local Similarity 27.5%; Pred. No. 8.1;
Matches 19; Conservative 11; Mismatches 20; Indels 19; Gaps 3;

QY 3 PHNKCLQSCNSERDSYRNQACHARCNLKVEKEBCE-----GEIPR 46
      | : : : | | | | | | | | | | | | | | | | | | | | | |
DB 211 PRDEQIKQAEAVES--KQAEATRLKIKITDREAEAEKRRADAKEDGKPPRPAEQP 268

QY 47 RPRPOHPE 55
      | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 APAPO-PEK 276

```

Search completed: March 1, 2001, 16:17:18  
Job time: 456 sec